DOT/FAA/AR-96/79

Office of Aviation Research
Washington, DC 20591

Test and Evaluation Plan for Threat Image Projection with the CTX 5000 Explosives Detection System

S. Cormier, Ph.D.
J. L. Fobes, Ph.D.

Aviation Security Human Factors
Program, AAR-510
William J. Hughes Technical Center
Atlantic City International Airport, NJ 08405

September 1996

Final

Document is available to the U.S. public through
the National Technical Information Service,
Springfield, Virginia 22161

U.S. Department of Transportation
William J. Hughes Technical Center

NOTICE

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT/FAA/AR-96/79		

4. Title and Subtitle TEST AND EVALUATION PLAN FOR THREAT IMAGE PROJECTION WITH THE CTX 5000 EXPLOSIVES DETECTION SYSTEM	5. Report Date September 1996 6. Performing Organization Code AAR-510
7. Author(s) S. Cormier, Ph.D. and J. Fobes, Ph.D.	8. Performing Organization Report No. DOT/FAA/AR-96/79
9. Performing Organization Name and Address U.S. Department of Transportation, Federal Aviation Administration William J. Hughes Technical Center Atlantic City International Airport, NJ 08405	10. Work Unit No. (TRAIS) 11. Contract or Grant No. DTFA03-94-C-00012
12. Sponsoring Agency Name and Address U.S. Department of Transportation, Federal Aviation Administration Associate Administrator of Civil Aviation Security, ACS-1 800 Independence Ave., SW Washington, D.C. 20591	13. Type of Report and Period Covered Final 14. Sponsoring Agency Code ACS-1

15. Supplementary Notes

Prepared by NYMA Inc., William Maguire and Jean Dunn, 500 Scarborough Drive, Egg Harbor Township, NJ

16. Abstract

The effectiveness of Threat Image Projection (TIP) as an element of the Screener Proficiency Evaluation and Reporting System (SPEARS) for checked baggage screening with the CTX 5000 is evaluated.

The CTX 5000 combines computed tomographic imaging and automated detection of explosives. This complex system requires that screeners learn to skillfully discriminate innocent bags that cause the system to alarm from genuine threats.

Testing is designed to determine the effectiveness of TIP in meeting the critical operational issues described in this plan. Specifically, is TIP effective in improving and maintaining screener performance in the detection of improvised explosive devices? Can the performance of individual screeners be monitored effectively with TIP? The testing will be conducted at airports where the CTX 5000 system is currently located.

17. Key Words	18. Distribution Statement
Threat Image Projection Computed Tomography Explosive Detection Systems	This document is available to the public through the National Technical Information Service, Springfield, Virginia, 22161

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	59	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

PREFACE

This test plan defines the investigation of the Threat Image Projection as an element of the Screener Proficiency Evaluation and Reporting System for checked baggage screening with the CTX 5000. The key FAA personnel supporting this testing are J. L. Fobes, Ph.D.; S. Cormier, Ph.D.; E. C. Neiderman, Ph.D.; J. M. Barrientos; and B. A. Klock with the Aviation Security Research and Development Division, Human Factors Program (AAR-510).

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

ACRONYMS AND ABBREVIATIONS

AvSec HF	Aviation Security Human Factors
CBT	Computer-Based Training
C_{fa}	Maximum Acceptable False Alarm Rate
C_h	Minimum Acceptable Target Detection Rate
COIC	Critical Operational Issues and Criteria
CT	Computed Tomography
CTI	Combined Threat Images
DOT	Department of Transportation
FAA	Federal Aviation Administration
EIC	Exploratory Issues and Criteria
GAO	General Accounting Office
HF	Human Factors
HFE	Human Factors Engineer
IED	Improvised Explosive Device
M_{FA}	Proportion of False Alarms Generated by the Machine
M_H	Proportion of Hits Generated By The Machine
MOE	Measure of Effectiveness
MOP	Measure of Performance
NAS	National Airspace System
N_D	The Total Number of Innocent Passengers/Bags Tested
N_T	The Total Number of Targets Tested
O_{FA}	Proportion of Operator False Alarms from Machine Generated Alarms
O_H	Proportion of Operator Hits from Machine Generated Alarms
OIC	Other Issues and Criteria
P_d	Probability of Detection
P_{fa}	Probability of False Alarm
ROC	Receiver Operating Characteristic
SDT	Signal Detection Theory
S_{fa}	Proportion of False Alarms Generated by the System
S_H	Proportion of Hits Generated by the System
SP	Scan Projection
SPEARS	Screener Proficiency Evaluation and Reporting System
SPI	Safe Passage International Ltd.
T&E	Test and Evaluation
TEP	Test and Evaluation Plan
TIP	Threat Image Projection

1. INTRODUCTION

1.1 General

The Federal Aviation Administration (FAA), working with the U.S. aviation industry, is developing new equipment and procedures to improve aviation security in the National Airspace System (NAS). Investigation of human factors is critical to the success of these efforts. The President's Commission on Aviation Security and the General Accounting Office (GAO) recognized this need and have recommended that there be a greater focus on human factors and training to complement advanced technologies.

The Screener Proficiency Evaluation and Reporting System (SPEARS) is being developed to improve and maintain the effectiveness of security screening personnel employed at airports. The SPEARS consists of two components: (a) an offline Computer-Based Training (CBT) system to teach screeners to detect various threat objects and (b) an online threat image projection (TIP) training and testing program to be employed at airport security checkpoints. This latter configuration is designed to further develop and maintain threat detection proficiency by insertion of simulated threat images into the normal flow of passenger bag images. The effectiveness of the CBT and TIP components will be addressed during separate test and evaluation (T&E) activities.

InVision's CTX 5000 scanner is a new technology application that combines computed tomography (CT) and automatic detection of explosives. This is a more complex system than baggage screeners have previously used. It demands that security personnel use a new set of skills to accomplish the task of screening for Improvised Explosive Devices (IED)s, including the ability to distinguish machine (CTX 5000) false alarms from real threats. CBT and TIP represent important training variables that need to be carefully evaluated.

This is a Test and Evaluation Plan (TEP) for TIP; CBT is addressed in a separate TEP. This document addresses the Critical Operational Issues and Criteria (COIC), Other Issues and Criteria (OIC), and the Exploratory Issues and Criteria (EIC) established by the FAA for the TIP component of SPEARS for IED screening of checked baggage with the CTX 5000 system.

1.2 Purpose

This Operational Test and Evaluation (OT&E) is being conducted to evaluate the ability of the SPEARS CTX TIP to improve, maintain, and monitor screener performance using the CTX 5000 to screen baggage for the presence of IEDs. Maintaining a workforce of adequately trained and performing X-ray screening personnel is critical to the mission of aviation security, both domestically and internationally. This TEP outlines the methods and procedures to be used in ensuring that TIP training and evaluation for operators of the CTX 5000 system meets the functional requirements established by the FAA as necessary to produce a capable workforce. TIP refers to the capability of inserting actual threat images into the stream of bag images in the operational environment. Specifically, the TEP outlines methods to determine the effects of TIP on screener performance in the operational environment and the utility of TIP as a means of monitoring performance and evaluating the effectiveness of CBT. The TIP evaluation will include the collection and analysis of empirical data on the operational and technical capabilities of the CTX 5000 system with TIP at two major U.S. airports, San Francisco (SFO) and Atlanta, and at the Manila airport.

1.3 Scope

The focus of this TEP is to evaluate the degree to which a TIP capability for the CTX system provides a means of evaluating screener accuracy and increases the ability of screeners to successfully resolve machine-generated threat alarms in checked baggage. Detection rates for IEDs (and explosives), using conventional X-ray screening, need to be improved (Fobes et al., 1995). The use of CT offers a number of potential advantages over X-ray screening. The volume images obtained contain much more information than the X-ray image, allowing objects to be viewed without clutter of overlapping images and with higher contrast. At the same time, the use of substantial computing capacity for CT image reconstruction in these scanners facilitates implementation of computer-aided IED detection. In computer-aided detection, the machine first analyzes the image for the presence of explosives. The human operator then decides which potential threat objects resemble IEDs.

The CTX OT&E will be conducted at the three CTX 5000 deployment sites. It will evaluate the ability of TIP-trained screeners to resolve threat alarms and detect IEDs in checked baggage using the CTX 5000 system.

1.4 Background

1.4.1 SPEARS Program

The FAA is responsible for ensuring the safety of air travel. Airports pose a challenge to security because they must be readily accessible to the public. To meet this challenge, the FAA has developed a security concept for airports. This involves a complex system of trained personnel, properly maintained and calibrated equipment, and appropriate procedures to provide multiple layers of security. This includes pre-board screening of carry-on and checked baggage and passengers.

A number of policies affect pre-board screening operations. Federal Aviation Regulation (FAR) Part 107, Airport Security, Section 107.20 states, "No person may enter a sterile area without submitting to the screening of his or her person and property in accordance with procedures being applied to control access to that area." FAR Section 108.9 and FAR Section 129.25 present screening policies for domestic and international airlines. Airlines may refuse to transport any person who does not consent to a search of his or her person and carry-on belongings. Checked baggage may also be examined for the presence of potential threats.

The threat to civil aviation security has changed in the last decade. Explosive device technology improvements have increased airliner vulnerability to bombings. Today, IEDs are less likely to be prefabricated. They can be assembled from a variety of materials and made to resemble innocent objects. Semtex and C-4, for example, can be molded into sheets and made to resemble books or radios. Terrorists have also learned to embed IEDs in electronic devices to make detection even more difficult. Timing devices have been miniaturized and digitized, compounding the difficulties of detection with conventional X-ray equipment.

For these reasons, the potential for complete aircraft destruction, with great loss of life and disruption of the NAS, has grown. This threat has increased the need for new airport security systems and operator training in these systems.

The SPEARS Program was initiated in response to a congressional mandate (Aviation Security Improvement Act of 1990, Public Law 101-604) directing the FAA to improve aviation security through the optimization of human factors elements in the U.S. airport security system. The evaluation of screener performance and effectiveness was emphasized to identify potential security improvements. An aviation security Department of Transportation (DOT) task force supported this emphasis by concluding that human performance was the critical element in the screening process.

The mandate directed that screeners be effectively trained to use threat detection equipment properly. The detection of explosive and incendiary devices was identified as critically important because of the potential for significant loss of life and aviation resources.

InVision was one company that responded to the SPEARS initiative by modifying system software so that TIP could be inserted into the normal stream of baggage.

1.4.2 Improvised Explosive Device Screening with the CTX 5000 System

InVision's CTX 5000 is an X-ray-based scanner that automatically screens for explosives. The scanner makes an X-ray and CT examination of each bag and computer software then analyzes the CT slices. If the software detects no threat, the bag is CLEARED and unloaded from the scanner. If a potential explosive threat is detected, the computer activates an alarm. At the workstation, the operator is provided with CT and X-ray images of the bag, an outline of the region identified as a potential threat, and information (such as density and mass) about the potential threat object. The operator, following the Alarm Resolution Procedures that are emphasized in training, examines the bag and determines whether the threat is real. If the operator determines that the bag is safe, the bag is CLEARED. If the operator cannot determine that the bag is safe, it is declared SUSPECT and additional security procedures are followed.

The CTX 5000 is a complex system. It requires that screeners learn to operate the controls and accurately interpret CT slices of checked baggage. The training component is critical to the success of this system. Cognitive and behavioral psychology provides information about how training should be organized, and this information has been incorporated into the design of the TEP.

1.4.3 Implications of the Training Literature

The training literature has implications for the design and implementation of TIP. In a review of training literature, Goldstein (1986) found that distributed practice for procedural skills, such as X-ray screening, provides the most advantageous results over time. Massed practice sessions tend to show better immediate training results and require less overall training time to achieve a minimum criterion. Massed, offline, instructional training is, therefore, a necessary component of screener training and is provided by CBT. For retention over extended periods, however, distributed or spaced training sessions will result in better overall performance. TIP, as a regular component of the screener's daily activities, should result in improved IED detection performance.

The schedule of TIP presentation is also expected to be an important variable, with variable rate presentation likely to be much more effective than fixed rate presentation (Schwartz, 1984).

Another factor critical to the acquisition and subsequent retention of job skills is the extent to which training is transferred to the operational environment. The TIP concept employs the principle of identical elements in that the training task takes place in the operational environment and the TIP targets are identical to real targets. Theory predicts that this training principle should result in a high level of transfer of TIP training to the operational environment.

1.4.4 Cognitive and Behavioral Analysis of IED Screening

Maintaining a high level of vigilance and performance in IED detection presents unique problems. The defining feature is that it is a discrimination task practiced under vigilance conditions, where the signal to be discriminated almost never occurs. This has a number of predictable effects.

Even well-trained screeners should, over time, show diminished ability to discriminate target from non-target. This is because, in the normal operational environment, the following occur:

a. The screener ceases to obtain information about threat images. While pattern recognition abilities are remarkably stable without reinforcing experience, the procedures that one should follow in doing a complex perceptual evaluation are highly susceptible to forgetting.

b. Any identified threat will usually be a false alarm, and the false alarms are likely to carry some aversive consequences. As a result, over time, screeners will be increasingly biased towards not identifying bags as threats (criterion shift).

A decline in ability to discriminate threats from non-threats is predicted to accompany a conservative shift in decision criterion in the normal course of screening. For this reason, TIP insertions into the luggage stream provide an important change in the contingencies of the task by providing threat information and response motivation.

Additional decline in performance is expected from vigilance decrement. Vigilance decrement in a monitoring task with infrequent targets, such as screening for explosives, is a decline over time in the probability of correctly reporting targets. Such decline can even be seen in the first hour. This decrement in a well-learned task is mainly due to a motivational component, attributable to an upward adjustment of the operator response criterion in response to a reduction in the perceived frequency (and therefore expectancy) of target events (Wickens, 1992). TIP, by increasing critical event frequency, should lessen vigilance decrement.

1.4.5 Signal Detection Theory and Analysis of the Task

Signal detection theory remains the method of choice when effects on performance can be either effects on sensitivity, decision criterion, or both. In the operational environment, it is generally only possible to measure a single operating point on the Receiver Operating Characteristic (ROC) curve. This presents special challenges to the evaluation of accuracy and criterion. The approach taken to system evaluation in this TEP is outlined in Appendix A. A description of signal detection theory is provided in Appendix B.

1.5 Functional Requirements

To justify the increased expense that CT screening of checked baggage represents, the system must be capable of detecting IEDs with a very high sensitivity. This must be done without unacceptably slowing the normal transport of baggage, increasing baggage delivery delays, or delaying airline takeoffs. The CTX 5000 system can detect explosives, however, it can only accomplish this by false alarming on a substantial number of bags, many more than can be practically searched by hand.

The system is specifically designed to work with human screeners who will examine X-ray and CT images of each bag that is alarmed. They will then determine which alarms should be CLEARED because the bag contains no threat and which alarms are SUSPECT requiring closer examination. In order for the system to succeed, the following must be true.

a. InVision modifies the equipment to allow for the insertion of Combined Threat Images (CTIs), with their corresponding alarms and related information. The modified system is capable of presenting CTIs on variable rate predetermined schedules. The mean rate and the range of rate variation can be controlled by a supervisor or other privileged user. The relative frequency of true and false CTI alarms can also be controlled. The system keeps track of the identity of the screener on duty, the time of CTI presentation, the type of alarm, the time at which the alarm was resolved, and the result of alarm resolution. All these data are kept in a log file. While waiting for InVision to complete the functional requirements to archive TIP presentations and their associated outcomes, Aviation Security Human Factors (AvSec HF) will create a text processing tool for extracting the information from the log file that is critical to this TEP. The system is also capable of providing the screeners with immediate feedback about the presence of CTIs following their resolution decisions. Screeners must be able to discriminate machine false alarms from genuine threat objects.

b. Alarming and alarm resolution must take place without causing significant delays in baggage processing.

c. Screeners must acquire the ability to resolve alarms in a relatively short time on the job and be able to sustain that ability throughout their tenure.

d. Supervisory personnel must be able to determine who is and who is not able to accomplish the task.

These four requirements are critical for the system to work in the operational environment. TIP plays a critical role in determining whether these requirements can be met. This is because TIP potentially provides a monitoring and evaluation capability in combination with a training and skill maintenance function. The ability of TIP to provide these different functions without causing slowdown in system throughput shall be investigated as described in this TEP.

1.6 System Description

1.6.1 CTX 5000 with TIP

InVision's CTX 5000 is an X-ray-based scanner that automatically screens for explosives (not explosive devices). The main unit consists of an X-ray, Scan Projection (SP) unit, and helical CT scanner. The operator sits at a workstation that has a console and display panel. The system is capable of presenting TIP.

1.7 Test Overview

1.7.1 Test Phases

Pilot testing. The first phase of testing will take place at SFO. It will involve the three to four CTX screeners who are currently on the job and have been performing these duties for at least 2 weeks before the pilot test. The screeners do not have CBT and were trained using the interim training syllabus developed by Lawrence Livermore National Laboratories and Invision under the direction of the FAA (Cormier & Fobes, 1996). The pilot testing will involve offline testing of screeners' alarm resolution abilities and a minimum two week exposure to TIP. If screeners remain available and conditions permit, screeners will continue with TIP beyond the two week period. Then, some longer term issues will be addressed: whether screeners can recognize repeated presentations of the same CTI and the rate of TIP presentation where it loses effectiveness as a motivator. These questions are important due to the modest size of the TIP library of images.

Operational Testing 1 (OT1). Operational testing will be restricted to screeners who have just had CBT. Because screeners are brought in as needed, different groups will begin CBT at different times. Operational testing for each group will continue as follows, but the start of operational testing will be staggered for different groups. A baseline offline test of alarm resolution abilities is followed by an 8-hour period of non-TIP work with the machine. This is followed by a 2-week period of carefully monitored TIP exposure. The OT1 period ends with a second offline alarm resolution test, followed by surveys of the operators and supervisors regarding the usability of TIP (Appendices D and E).

Operational Testing 2 (OT2). The second operational testing period begins immediately after the end of OT1. The rate of TIP presentation and the intervals between repeated presentations of the same TIP image will be adjusted to a rate consistent with long term maintenance of TIP benefits, as determined by pilot results. Screener performance with reduced rates and varied repetition intervals of TIP presentation will be tracked for 2 months or until the employee terminates, if this occurs before two months have passed.

In-House Technical Testing. A variety of tests will be carried out at the AvSec facilities and in the field to evaluate aspects of the OICs (see section 1.8.2) amenable to laboratory and technical testing. These will be completed at the same time as the pilot testing.

1.7.2 Testing Limitations

The number of screeners and sites is relatively small and generalization to other sites must be done carefully. It is not possible to try a wide range of TIP presentation rates to determine the optimal rate to maintain or improve performance. Instead, a rate of TIP presentation that is practically feasible, and produces tangible benefits will be sought and used even though there may be more optimum rates discoverable through systematic parametric research.

Offline testing is necessary to determine the performance levels of individual screeners. In the operational setting, screeners occasionally work jointly with a supervisor to resolve alarms. Thus, individual performance can be confounded in some instances by a group decision or the supervisor overruling or influencing a screener decision. In addition, an offline test is the only way in which a standardized set of images can be administered to all screeners across sites. Since TIP involves the presentation of CTIs amid actual baggage streams, there is no control over the number and types of machine alarms that the screener will be exposed to during TIP. In addition, because of the nature of the operational environment, it is not possible to guarantee that a particular screener will be on station at the time that a particular CTI appears. It is also impossible to know whether the screener's decision will be affected or changed by the (intermittent) presence of the supervisor. Offline testing, however, is characterized by a particularly high target incidence rate that can have some effect on screeners' decision criteria.

1.7.3 Test and Evaluation Milestones

Table 1 shows the milestones for planning and reporting the T&E.

Table 1. Test and Evaluation Milestones

MILESTONE	DATE	RESPONSIBLE ORGANIZATION
TEP Finalized	9/15/96	AvSec HF
Pilot Testing	8/6/96-10/1/96	AvSec HF
Operational Testing Phase 1	9/7/96 - 11/7/96	AvSec HF
Operational Testing Phase 2	10/7/96 - 1/7/97	AvSec HF
In House Testing	8/1/96 - 10/1/96	AvSec HF
Draft T&E Report	3/97	AvSec HF
Final T&E Report	5/97	AvSec HF

1.8 Overview of Issues and Criteria

Issues and criteria are divided into three classes for the assessment of TIP against the functional requirements: COIC, OIC, and EIC. These issues involve substantially different requirements and investigative methods. The COIC are evaluated in operational testing, the EIC in pilot testing, and the OIC generally by in-house testing.

1.8.1 Critical Operational Issues and Criteria

The COIC are those issues and criteria necessary to evaluate the TIP operational requirements. Each issue is analyzed in terms of one or more criteria by which the system is judged. Each criterion leads to one or more Measures of Performance (MOPs) and Measures of Effectiveness (MOEs).

The COIC broadly fall into two categories: issues related to TIP as a measurement of performance and issues related to TIP as a vehicle for training and maintaining vigilance.

TIP itself is potentially a useful means of training screeners to recognize IEDs and motivating screeners to be alert for them (Issue 1). In practice, it may be difficult to separate changes in performance due to learning from changes in performance due to motivation. While we will track both long- and short-term performance changes in the operational testing, we may not be able to specify the relative roles of learning and motivation in those performance changes. Issue 2, Usability, concerns the broad need to have a TIP capability that is reliable and can be operated by the personnel in place. Issue 3 concerns the evaluation of offline tests as tests of individual proficiency.

1.8.2 Other Issues and Criteria

The OIC are those issues and criteria that are supplementary, more specific, or technical in nature. They include system customization, screener capabilities reporting, downloading capability, feedback, and security. The OIC will be investigated using structured protocols, operational, and in house testing of system features. Checklists that are used will make use of the Human Factors Deficiencies Rating Scale (Appendix C).

1.8.3 Exploratory Issues and Criteria

The EIC (pilot testing) will be concerned with three main exploratory issues: determining the proper presentation rates and repeat rates for TIP during the OTE, evaluating issues concerned with offline testing during the OTE, and determining that data collection procedures are reliable and adequate.

For the future, there are a limited number of CTIs available for TIP. This means that there is a limited period within which unique CTIs can be presented without repetition. This period is naturally a function of the presentation rate. Issues 11, 12, and 13 are concerned with the minimum effective TIP presentation rate, the period over which previously presented images are remembered, and the possible use of a low presentation rate to maintain TIP benefits over the long term. These issues all are directed to the most effective use of a limited number of images.

Operational testing involves pre- and post-TIP offline tests. Differences in performance between the first and second offline test will be used to interpret issues such as the effectiveness of TIP as a training vehicle. Validity of these MOEs depends upon the two offline tests being approximately equal in difficulty. Issue 11 describes pilot procedures to attain this end.

A second important component of the offline test scores is the degree to which the tests are sensitive enough to reflect real performance differences. The utility of power analysis techniques to specify the minimum number of test items for statistical analysis will be investigated.

1.9 CT TIP Critical Operational Issues And Criteria

1.9.1 Issue 1 - Effectiveness of TIP in Maintaining Vigilance

Criterion 1-1 Does vigilance change over the course of weeks of TIP exposure?

MOP 1-1-1 Hit rates, false alarm rates, and alarm resolution times measured each day.

MOE 1-1-1 Screener performance is maintained throughout the test period.

Criterion 1-2 The presence of TIP does not substantially increase alarm resolution time or degrade system throughput.

MOP 1-2-1 Alarm Resolution time during TIP and during the pre-TIP period.

MOE 1-2-1 Mean alarm resolution time is not greater under TIP.

MOP 1-2-2 Average bag delay during TIP and during the pre-TIP period.

MOE 1-2-2 Changes in average bag delay with TIP are not operationally significant.

Does the effect of TIP on screener performance apply broadly to all applicable categories of explosives?

Criterion 1-3 Performance is acceptable for all alarm classes.

MOP 1-3-1 Machine, screener's (machine dependent) hit and false alarm rates, and overall system hit and false alarm rates.

MOE 1-3-1 Screener's (machine dependent) hit rate is greater than the false alarm rate for every alarm class subset.

The minimum standard for TIP effectiveness is that performance in the second offline test be at least roughly equivalent to performance in the first offline test. Video records obtained during CBT evaluation will make it possible to estimate the temporal distribution of bag arrivals to the scanner during quiet and busy periods and the average bag delay at the scanner. From the computer records, the average time to scan a cleared bag or to scan and resolve an ALARMED bag can be estimated for pre-TIP and TIP screening. A simple single queue model, where the machine alarm rate, screening time, and alarm resolution time during TIP and pre-TIP are the important parameters, will be used to evaluate the performance criteria.

There are not enough threat images and false alarm images to insure that each screener will receive sufficient numbers for each alarm class. Therefore, the only measures robust enough to be examined for individual alarm classes are pooled measures of performance over the full TIP duration. The minimum screener performance expected for each alarm class is that their machine contingent decisions discriminate within each category of alarm, at better than chance expectancy.

Criterion 1-4 Does the vigilance of screeners change over the course of the daily shift?

MOP 1-4-1 Hit rates, false alarm rates, and alarm resolution times measured for early vs. late parts of the shift.

MOE 1-4-1 Screener performance is maintained throughout the shift.

Each day early in the shift, vs. late in the shift, alarms will be recorded by noting the median number of alarms for a screener's shift. The alarms will be split into those that preceded and those that followed the median alarm. This information will be recorded for each screener each day, and will continue to be collected over the course of months in the extended TIP period. Changes in vigilance will be recorded by examining the pattern of early and late hits and false alarms as well as the pattern of alarm resolution times, over the course of days.

1.9.2 Issue 2 - TIP Usability

Are there any software or hardware factors or procedures that decrease TIP effectiveness?

Criterion 2-1 Investigative in nature.

MOP 2-1-1 Deficiencies noted from human factors design standards.

MOE 2-1-1 No severe deficiencies are found.

Can images be inserted in uncued fashion into the baggage stream?

Criterion 2-2 The TIP images differ from normal bag images only in that they include a test object.

MOP 2-2-1 Any perceptible artifacts or distinctive features associated with the CTIs.

MOE 2-2-1 No distinctive features are found that are not due to the presence of the IED simulants.

MOP 2-2-2 Screeners' reports about the presence of artifacts.

MOE 2-2-2 No reports of image artifacts.

MOP 2-2-3 Screeners' performance for naturally occurring machine false alarms and TIP false alarms.

MOE 2-2-3 Screeners are not more likely to call TIP false alarms SUSPECT than they are naturally occurring false alarms.

Criterion 2-3 Can supervisors use the system as intended?

MOP 2-3-1 Difficulties reported in use by supervisory personnel.

MOE 2-3-1 Supervisory personnel report they can carry out all the required functions using the system.

Usability evaluations will consist of feedback from the operators and supervisors and evaluation of usability by HFEs. Screeners will be given the Screeners' Survey on TIP Usability (Appendix D) and Supervisors, the TIP Usability Supervisors Survey (Appendix E). Screeners will be asked to provide some basic demographic information (Appendix F). Informed consent (Appendix G) will be obtained from screeners before these questionnaires are administered. The TIP Usability Human Factors Engineering Checklist uses scales based upon the Human Factors Deficiency Rating Scale (Appendix C) and will be utilized by HFEs in their evaluation of usability. This checklist is an adaptation of the Guidelines for the Design of User Interface Software (Smith & Mosier, 1986) and MIL-STD 1472D and found in Appendix H.

Screener performance will be compared between passenger bags that trigger machine alarms and CTI false alarms inserted into the baggage flow. Differences in performance between these two classes of images would be suspicious and require investigating. HFEs will evaluate whether extraneous cues may signal the presence of a CTI.

1.9.3 Issue 3 - TIP as a Measure of Individual Performance

Are individual performance measures obtained in offline testing valid?

Criterion 3-1 Do the images used in offline testing show item validity?

MOP 3-1-1 Number of test images that show 0 or 100% alarm resolution performance averaged across screeners.

MOE 3-1-1 Most images are in the middle range of item difficulty.

This issue will be examined both in the pilot and operational testing. The quality of individual test items will be determined following operational testing by summing performance on each item over all screeners.

1.10 CT TIP Other Issues and Criteria

1.10.1 Issue 4 - Customization

Can TIP rate be controlled?

Criterion 4-1 CTI presentation rates can be controlled.

MOP 4-1-1 Deficiencies noted in control of TIP presentation.

MOE 4-1-1 No severe deficiencies noted in control of presentation.

The operational situation does not favor customization of parameters for individual screeners for reasons noted above. However, the rate of TIP presentation is one factor that can be usefully customized to some extent. The evaluation of TIP customization will take place during in-house testing and in the operational environment where the ability of supervisory personnel to control TIP will be evaluated using Appendix I.

1.10.2 Issue 5 - Feedback

Can TIP feedback presentation and timing be controlled?

Criterion 5-1 CTI feedback can be controlled.

MOP 5-1-1 Deficiencies noted in control of feedback.

MOE 5-1-1 No severe deficiencies noted in control of feedback.

The evaluation of TIP feedback will take place during in house testing, and in the operational environment using Appendix J.

1.10.3 Issue 6 - Screener Capability Summaries

Are useful training reports prepared?

Criterion 6-1 TIP training reports contain individual and cumulative descriptive statistics that adequately summarize screeners' performance as individuals and as groups. These reports give both relative and absolute measures of performance.

MOP 6-1-1 Deficiencies found in TIP training reports.

MOE 6-1-1 No severe deficiencies found in TIP reports.

The evaluation of screener capability reports will be carried out by HFEs in house and in the field using the checklist in Appendix K.

1.10.4 Issue 7 -Downloading

Can the equipment send image displays to remote computers?

Criterion 7-1 CTIs can be sent to remote sites.

MOP 7-1-1 Deficiencies found in transmitting CTIs to remote sites.

MOE 7-1-1 No severe deficiencies found in transmitting CTIs.

The evaluation of downloading will proceed using the checklist in Appendix L.

1.10.5 Issue 8 - Security

Is access restricted?

Criterion 8-1 Only authorized personnel can access particular aspects of the system.

MOP 8-1-1 Deficiencies noted in system security.

MOE 8-1-1 No severe deficiencies in security.

During the pre-training evaluation, HFEs will test system security. Deficiencies will be noted by use of the, Security Access Control Checklist (Appendix M).

1.11 Exploratory Issues

These are issues that will be explored in pilot testing. It is understood that these issues cannot be decisively settled with the time and the number of screeners available.

1.11.1 Issue 9 - Pre-testing Operational Evaluation Procedures

Criterion 9-1 The maximum practical TIP presentation rate that maintains screener performance.

MOP 9-1-1 Screener performance over TIP trials.

MOE 9-1-1 TIP performance does not decline over days.

Criterion 9-2 Image sets for pre- and post-TIP offline testing are of equal difficulty.

MOP 9-2-1 Performance for each image presented in offline tests averaged across screeners.

MOE 9-2-1 Image sets can be rearranged to equate for difficulty.

Criterion 9-3 The size of offline testing image sets is sufficient to test the key operational issues.

MOP 9-3-1 Power analysis of offline test performance during the pilot.

MOE 9-3-1 The likelihood of a type II error is less than 35 percent.

Criterion 9-4 The data collection procedures are effective.

MOP 9-4-1 Deficiencies in data collection noted during the pilot testing.

MOE 9-4-1 No severe deficiencies are found.

The pilot testing will be done with experienced CTX operators in SFO. Four screeners will be given the offline tests, go through a period of TIP training and, if available, a period of extended TIP at reduced presentation rates. One important focus of this activity will be to make sure that all aspects of TIP evaluation are adequate to answer the COICs. Another focus will be to make sure that the rate of TIP presentation chosen is efficacious. These issues will be evaluated by HFEs who are directing the testing at SFO. The details of this evaluation are discussed in section 2.4.

1.11.2 Issue 10 - Reduced TIP

Does performance deteriorate with TIP reduction?

Criterion 10-1 Performance does not decline when TIP is reduced over the long term.

MOP 10-1-1 Operator hit and false alarm rates plotted on a daily basis for TIP and TIP reduction periods.

MOE 10-1-1 Average daily hit and false alarm rates are not affected by reduced TIP.

Operator performance during TIP is recorded on a daily basis. During extended TIP, the schedules of TIP presentation will be reduced by half. Performance over the period of extended TIP will be compared with the earlier TIP period.

1.11.3 Issue 11 - Repeated Presentations

Does repeated CTI presentation reduce the total number of CTIs required for long-term TIP?

Criterion 11-1 There is not substantial remembering of old CTIs after three weeks.

MOP 11-1-1 The ability to discriminate previously presented CTIs after three weeks.

MOE 11-1-1 After three weeks, CTIs are not recognized as previously presented.

In the fourth week of the extended TIP period, the screeners will be shown a set of CTIs. Some of these will have been presented during week 1 of TIP training and some never seen before. The screeners will be asked simply to say whether they had ever been exposed to this CTI before and how confident they were in their judgment. A five point likert scale and the data sheet shown in Appendix N will be used.

2. PILOT TESTING

The pilot testing is designed to field test the data collection procedures and to obtain data relevant to the investigational issues listed.

2.1 Subjects

The subjects will be four screeners at SFO who have worked with the CTX 5000 for some months.

2.2 Equipment

The pilot test will be run on the CTX 5000 with TIP capability in the checked baggage handling area. Offline testing will be conducted on the CTX 5000 during periods when it is not processing baggage.

2.3 Data Collection Procedures

On the first day of pilot testing, all the screeners will be given the offline test for the first time. This test consists of the screener resolving 30 alarms, 15 true alarms and 15 false alarms. These alarms are drawn from the CTX library of CTIs, which are resident on the CTX hard disk. Thirty true and 30 false alarms will be pre-selected and divided into two groups (A, B) based on an SME's attempt to equate the sets for difficulty. Two of the screeners will receive A during the first offline test, and two will receive B. The next day the same screeners will be given a second offline test, and those who received set A initially will be given set B and vice versa.

On days 3-12, the screeners will undergo TIP training in the operational environment. The TIP presentation rate initially tried will be 1 CTI / 75 bags (range 1 /50 - 1/ 100 bags). Eighty percent of CTIs will be IEDs. Log files will record important data during the days of TIP presentation.

In the third week of TIP training, the beginning of extended TIP, the rate of TIP presentation will be lowered. This will continue for one month.

In the fourth week, screeners will be shown a set of images offline. These will be a set of 16 images, 8 of which were presented during the first week of TIP and 8 of which were never presented before to the screeners. Using a 5-point scale, the screeners will rate each image for their confidence that they have seen it before (see Appendix N).

2.4 Data Analysis

During testing, a log file is continuously updated. It contains information critical to subsequent data analyses. A text processing tool will be developed and used to generate session records for each log file. The session record will record the number of bags processed and the number of machine alarms per shift. Then for each machine alarm, it will record the following information:

1) Operator I.D.
2) Operator login time.
3) Bag I.D.
4) CTI Status: Threat? False Alarm?
5) Alarm class.
6) Screener's decision.
7) Time of initial bag image presentation.
8) Time of potential threat status indication.
9) Time of screener's alarm resolution.

From these records, all performance measures (hit rates, false alarm rates, alarm resolution times, and derived measures) will be calculated. Detailed discussions of these measures are found in Appendices A and B.

2.4.1 Pre-test of the Operational Evaluation Procedures

Since the pilot closely mimics the procedures that will be used in the operational test, problems that arise will be noted and changes in procedures made to eliminate methodological problems.

In particular, the early days of TIP training will be closely observed to determine whether TIP presentation rate appears to be effective. Adjustments will be made on the basis of those early observations.

In addition, a number of specific questions will be asked.

1) Are the image sets used for the first and second offline test matched for difficulty?

   Hit rates and false alarm rates for image set A and B should be equal when summed across screeners. Images should be exchanged between the sets to equalize these rates. Item analysis will be performed to evaluate the quality of individual test images.

2) Is the offline test set of 30 images large enough to evaluate CBT and TIP?

   Following the offline test, a power analysis (Cohen, 1988) will be performed to determine the sample size needed to detect expected changes in screener performance.

3) Can the offline testing be accomplished in a single session without significant fatigue?

   Alarm resolution time is recorded on a case by case basis for the offline tests. There should not be significant increases in alarm resolution time at the end of the session.

4) Do TIP presentation rates closely follow the preset parameters?

   Since the machine has some leeway in inserting a TIP image, it is important to ascertain that the preset presentation rate corresponds to the actual presentation rate.

2.4.2 Performance Under Reduced TIP Conditions

It is not known what rate of TIP presentation is needed to preserve the TIP performance objectives. An initial rate will be set based on the average rate of baggage flow at the SFO site that seems adequate and reasonable. This initial rate may be changed if it appears to be ineffective. Once a stable rate is found, investigations of the extent to which that rate can be reduced without affecting performance will be examined. Machine dependent operator hit and false alarm rates for TIP and extended TIP will be compared in the manner described in Appendix A.

2.4.3 The Effect of Repeat Presentations

Following the memory tests given in the second week of extended TIP, t-tests will compare ratings of old images to new images to see if there is any detectable memory of the images.

3. OPERATIONAL TEST AND EVALUATION

3.1 Subjects

The exact number of subjects will depend upon the availability and suitability of candidate screeners at the airport test sites. All screeners will have undergone CBT before introduction of TIP. As new screeners receive CBT, they will be introduced to the TIP protocols and enlisted into the study. We will use all new trainees that are produced in the period from 9/1/96 to 11/15/96. We anticipate that the number of subjects will range from 10 to 15.

3.2 Equipment

TIP trials will take place on the CTX 5000 system installed at the baggage processing areas at the three test sites. The scanner will be modified so that CTIs can be inserted into the baggage flow. Information about time and type of threat alarms, the screener's decision in each case, and the time of the decision will be recorded. Immediate feedback will be provided to the operator for all identified CTI targets.

3.3 Data Collection Procedures

Following CBT, each screener will spend at least 8 hours screening with the CTX under non-TIP conditions to familiarize screeners with the machine and procedures before measuring TIP effects. False alarm rates and alarm resolution times will be recorded during the pre-TIP period.

The first offline test immediately follows the non-TIP period. This test involves the presentation of 15 true and 15 false CTX alarms for resolution. All screeners will be tested with the same images.

TIP training begins immediately after the first offline test. The rate of TIP presentation will be the optimal rate identified during the pilot test. All TIP presentations are made on a variable ratio schedule. TIP training will last until at least 50 CTIs are presented. Following TIP training, the second offline test will be given with presentation of 15 true and 15 false CTX alarms for resolution. Each screener receives the same images, and none of the images will have been presented during OT1. Hit, false alarm rates, and alarm resolution times are recorded.

The number of threat and false alarm CTIs needed to accomplish the evaluation are listed in Table 2.

Table 2. CTI Requirements for TEP

	Day 1-2 Baseline NTIP	Day 3 Offline Testing	Day 4-14 TIP Training	Day 15 Offline Testing	Day 16-76 Extended TIP	Total Images
Hits	0	15	60	15	60	150
False Alarms	0	15	12	15	12	54

3.4 Data Analysis Procedures

The data analysis involves calculating hits, false alarm rates, and alarm resolution times for all measurable system components as illustrated in Table 3. The COIC are evaluated by comparison of various system components to each other or during different operational testing phases. The logic of these analyses, as applied to measuring system, machine, and operator performance, is explained in Appendix A. Comparison of alarm resolution times in different periods will be done by analysis of variance.

Table 3. Analytical Activities

Title/duration	Description	Measures Recorded
Phase 1		
Pre-Tip Day 1-2	Screeners work without TIP.	False Alarms, Alarm Resolution Times
Offline Test Day 3	CBT evaluated and baseline established.	Hits, False Alarms, Alarm Resolution Times
TIP testing Days 4 - 14	Screeners work with TIP on 5 per day schedule.	Hits, False Alarms, Alarm Resolution Times by day and position in shift
Offline Test Day 15	TIP effectiveness evaluated.	Hits, False Alarms Alarm Resolution Times
Usability Survey Day 15	Screeners and supervisors complete Usability Surveys.	Supervisor Usability Survey Operators Usability Survey
Phase 2		
Extended TIP	TIP on a reduced 2 per day schedule.	Hits, False Alarms, Alarm Resolution Times by day and position in shift

4. REFERENCES

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cormier, S. & Fobes, J. (1996). *Test and evaluation plan for computer based training with the CTX-5000 explosives detection system* (DOT/FAA/AR-96/67). Atlantic City, NJ: FAA Technical Center.

Fobes, J. L., McAnulty, D. M., Klock, B. A., Barrientos, J. M., Weitzman, D. O., Fischer, D. S., Malone, R. L., Janowitz, J., & Dolan, N. (1995). *Test and evaluation report for the screener proficiency evaluation and reporting system (SPEARS) CBT* (DOT/FAA/CT-95/10). Atlantic City, NJ: FAA Technical Center.

Goldstein, I. L., (1986). *Training in organizations: Needs assessment, development and evaluation* (2nd edition). Pacific Grove, CA: Brooks/Cole.

Schwartz, B. (1984). *Psychology of learning and behavior* (2nd ed.). New York: Norton.

Wickens, C. (1992). *Engineering psychology and human performance* (2nd ed.). New York: Harper-Collins.

APPENDIX A

TESTING HYPOTHESES ABOUT SYSTEM PERFORMANCE

System Overview

We define the operator(s) and the CTX 5000 working together as the explosives detection system (EDS). The performance of the EDS depends critically on the performance of the individual parts, machine, and operator.

The relationship between machine hits and false alarms and ensuing screener hits, misses, false alarms, and correct rejections is depicted in Figure A-1. When the machine alarms a bag and presents the images to a screener for final decision, either the bag actually contained an IED (machine hit) or it did not (machine false alarm). For a machine hit, the screener either confirms that an IED is present in the bag (screener hit) or incorrectly rejects the determination made by the machine (screener miss). The screener probability of detection, which we call the machine contingent screener hit rate (O_H) is calculated as the ratio of screener hits to machine hits. The screener's performance is machine contingent, meaning the screeners' hit and false alarm rates will be contingent on the particular group of bags that the machine alarms.

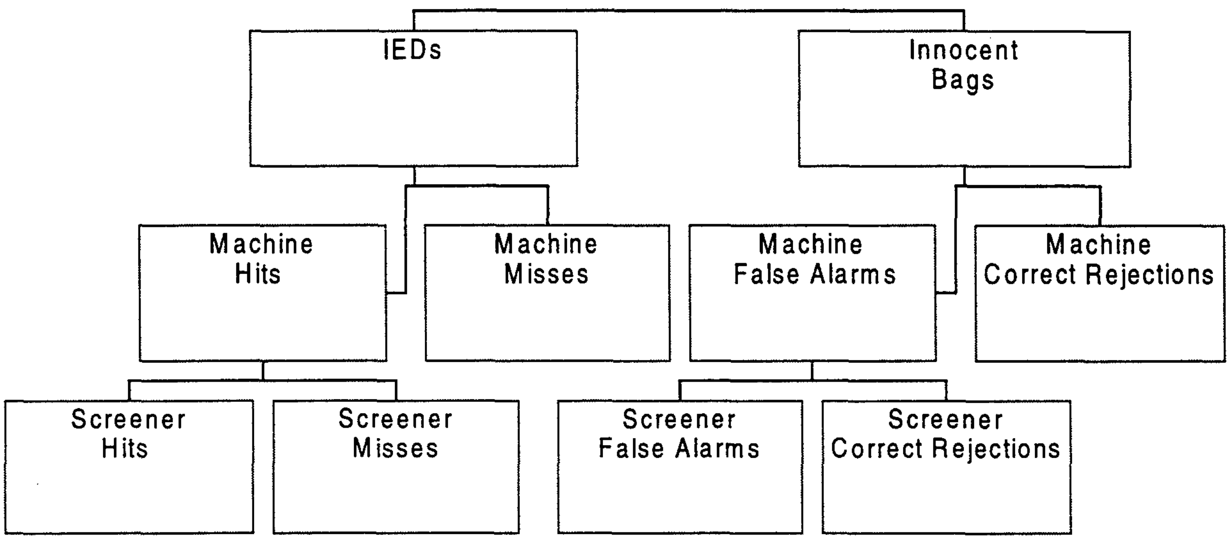

Figure A-1. Decision flow diagram for CTX 5000 EDS

For machine false alarms, the screener either incorrectly accepts the determination made by the machine (screener false alarm) or correctly rejects the determination made by the machine, (screener correct rejection). The machine contingent screener false alarm rate (O_{FA}) is the ratio of screener false alarms to machine false alarms.

To calculate true system parameters, as shown in the Table A-1, both machine and screener hit and false alarm rates must be known.

Measure	Definition	Formula
N_T	Number of IED bags	
N_D	Number of innocent bags	
N_A	Number of IED bags that the machine alarms	
N_{AD}	Number of innocent bags that the machine alarms	
M_H	The machine hit rate	N_A / N_T
M_{FA}	The machine false alarm rate	N_{AD} / N_D
O_H	The screener hit rate	N_H / N_A, S_H / M_H
O_{FA}	The screener false alarm rate	N_{FA} / N_{AD}, S_{FA} / M_{FA}
N_H	Number of IED bags the system (M &O) detects	
N_{FA}	Number of innocent bags the system calls suspect	
S_H	The system hit rate	N_H / N_T
S_{FA}	The system false alarm rate	N_{FA} / N_D
C_H	Minimum acceptable hit rate	
C_{FA}	Minimum acceptable false alarm rate	

Table A-1. Symbols for system parameters and their formulas

The ideal system descriptions are not in terms of hits and false alarms but in terms of ROC parameters that describe sensitivity and decision criterion separately. This is because system performance depends jointly on the sensitivity of system components to explosives and the decision criterion adapted (particularly by the operators, but also analogous criterion choices may have been made in the design of the system). The data collection methods, however, only permit the collection of a single system operating point (hit rate and false alarm rate pair). The analysis below begins with analysis of hit and false alarm rates. The use of signal detection analysis to clarify ambiguous situations is discussed in Appendix B.

Testing Overall System Effectiveness

The critical question for system effectiveness is whether the system performs to an acceptable standard. We wish to test the hypothesis that

$S_H >$ Minimum acceptable target detection rate C_H.

$S_{FA} <$ Maximum acceptable false alarm rate C_{FA}.

The null hypothesis to be tested is $S_H < C_H \vee S_{FA} > C_{FA}$, a disjunctive combination of simple hypotheses about hit and false alarm rates. The specific test to be used depends critically on the sample sizes, either an exact binomial test of proportions or a test based upon a normal approximation to the binomial distribution. However, the logic is always the same. Determine the probability that the data are consistent with $S_H < C_H$, **p** and the probability that the data are consistent with $S_{FA} > C_{FA}$, p_2. The alpha level for the test of the disjunction is $p = p_1 + p_2 - p_1 * p_2$. This means that even if both simple hypotheses are significant at $p<.05$, the disjunctive hypothesis still might be rejected at $p<.05$. The simple hypotheses should be tested at $p<.02$ approximately.

If the hit rate is acceptable but false alarm rate unacceptable, or vice versa, one would want to estimate ROC parameters as shown below. Despite the great uncertainty in projecting one operating point from another, there are some minimum predictions that can be made.

Testing The Effectiveness Of The Operator Within The System

Is the operator sensitive to the difference between targets and false alarms?

The most fundamental hypothesis to test about the operator is whether their contribution to the system is more than a simple shift in decision criterion. Specifically, we can test the hypothesis that the machine contingent operator hit rate is greater than the machine contingent operator false alarm rate, which would mean that A_z is greater than .5.

The null hypothesis: $O_H <= O_{FA}$

This is tested by a chi square test if all cells contain at least 5 entries, otherwise Fisher's Exact Test.

APPENDIX B

SIGNAL DETECTION THEORY

The Signal Detection Theory (SDT) Paradigm

The SDT is based on the concept that perception is affected by the expectations of the observer and the payoffs associated with the consequences of judgments and the physical input to the receptors (Foley & Morey, 1987). In other words, perception is determined by the interaction of the physical parameters of the stimulus with the subjective control of the perceptual mechanisms by the observer. The theory of signal detection is applicable to situations in which there are two discrete states (e.g., signal and noise) that cannot be easily discriminated (Wickens, 1992).

The Concept of Noise

A central concept in SDT is that, in any situation, there is noise that can interfere with the detection of a signal. This noise can be generated externally (e.g., noises in a factory other than a warning buzzer) and/or by the observer (e.g., variations in neural activity). This noise varies over time, thus forming a distribution of intensity from high to low. The shape of this distribution is assumed to be normal (i.e., bell shaped). When a signal occurs, its intensity is added to that of the background noise. At any given time, a person needs to decide if the sensory input (i.e., what the person senses) consists of only noise or noise plus the signal (Sanders & McCormick, 1987).

Possible Outcomes

The IED detection task can be looked at with this model. The screener must detect an environmental event or signal (i.e., a threat object). Based on SDT, there are two responses that represent the screener's detection performance:

YES (a threat object was present), or

NO (a threat object was not present).

There are also two signal presentation states:

SIGNAL (a threat object was present), or

NOISE (a threat object was not present).

The combination of the screener responses and the signal state produces a 2 x 2 matrix (refer to Figure B-1) comprised of four quadrants. These quadrants are labeled hits, misses, false alarms, and correct rejections.

Signal Presentation State

		SIGNAL	NOISE
Screener Response	**YES**	HIT	FALSE ALARM
	NO	MISS	CORRECT REJECTION

Figure B-1. Matrix of Screener Response and Signal Presentation State

<u>The Concept of Response Criterion</u>

One of the major contributions of SDT to the understanding of the detection process was the recognition that people set a criterion along a hypothetical continuum of sensory activity and that this activity is the basis upon which a person makes a decision. The position of the criterion along the continuum determines the relative probabilities of the four outcomes listed above. The best way to conceptualize this is in Figure B-2. Shown are the hypothetical distributions of sensory activity, when only noise is present and when a signal is added to the noise. Notice that the two distributions overlap. That is, sometimes conditions are such that the level of noise will exceed the level of noise plus signal.

As indicated by Wickens (1992), the SDT paradigm assumes that operators perform two stages of information processing in all detection tasks: (1) sensory evidence is aggregated concerning the presence or absence of the signal, and (2) a decision is made about whether this evidence indicates a signal. According to SDT, external stimuli generate neural activity. On average, there will be more neural activity when a signal is present than when it is absent. This neural activity increases with stimulus magnitude but there are external and internal sources of noise that make the activity level indeterminate. Therefore, the observer chooses a threshold. If the level of neural activity exceeds a critical threshold, X, the operator decides "yes." If it does not, the operator decides "no." This is illustrated in Figure B-2.

SDT postulates that a person sets the criterion level such that whenever the level of sensory activity exceeds that criterion, the person will say there is a signal present. When the activity level is below the criterion, the person will say there is no signal. Figure B-2 also shows four areas corresponding to hits, false alarms, misses, and correct rejections based on the criteria shown in the figure.

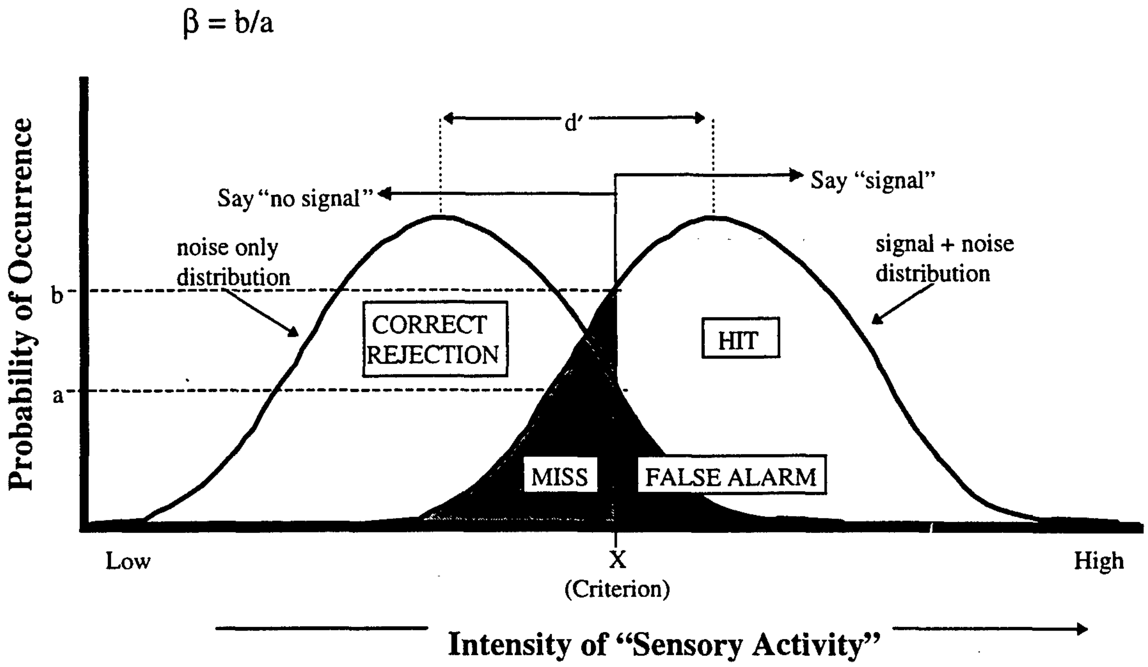

Figure B-2. Key Concepts of Signal Detection Theory.

Related to the position of the criterion is the operator response criterion quantity (i.e., beta). Beta is the ratio (signal to noise) of the height of the two curves in Figure B-2 at a given criterion point. As the criterion is shifted to the right, the response criterion value increases and the person will say signal less often and hence will have fewer hits, but also fewer false alarms. A large value of Beta represents a conservative judgement. A small value of Beta will have more hits, but also more false alarms and represent a riskier judgement (Sanders & McCormick, 1987).

For the IED detection tests used in the demonstration testing:

a. A Hit will be recorded when a screener correctly identifies a threat image as a threat.

b. A False Alarm will be recorded when a screener incorrectly identifies an image of an innocent bag as a threat reports.

c. A Miss will be recorded when a screener clears a threat image.

d. A Correct Rejection will be recorded when a screener clears an innocent bag image.

Figure B-2 illustrates the key concepts of SDT. Shown are the two hypothetical distributions of internal sensory activity, one generated by noise alone and the other generated by signal plus noise. The probabilities of four possible outcomes are depicted as the respective areas under the curves, based on the setting of a criterion at X. Here, d′ is a measure of sensitivity, and β is a measure of response bias. The letters *a* and *b* correspond to the height of the signal-plus-noise and noise-only distributions at the criterion.

B.5 Influencing Response Criterion

SDT postulates two variables that influence the setting of the criterion: (1) the likelihood of observing a signal, and (2) the costs and benefits associated with the four possible outcomes. Consider the first. If field intelligence indicated that a terrorist had threatened a flight, thus increasing the likelihood of a threat object, the screener would be more likely to regard an alarmed bag as a threat object. That is, as the probability of a signal increases, the response criterion is lowered (reducing β), so that anything remotely suggesting a signal (a threat object in our example) might be called a signal (Sanders & McCormick, 1987).

With regard to the second factor, costs and benefits, again consider the tasks of the airport screener. What is the cost of a false alarm (saying there may be a threat object when there is not)? The bag is hand searched and throughput at the checkpoint is decreased. What is the cost of a miss (saying that no threat object exists when there is one)? The threat object may get on to the aircraft and an act of terrorism may occur. Under these circumstances, the screener may set a low criterion and be more willing to call a suspicious object a potential threat. But what if the bag has to pass through three different checkpoints with different types of sophisticated detection equipment? In this case, the screener would set a more conservative criterion and would be less likely to call it a threat object.

Concept of Sensitivity

In addition to the concept of response criterion, SDT also includes a measure of the person's sensitivity, that is, the acuteness of the senses. Further, SDT postulates that the response criterion and sensitivity are independent of each other. In SDT terms, sensitivity is measured by the degree of separation between the two distributions shown in Figure B-2. The sensitivity measure is called d′ and corresponds to the separation of the two distributions expressed in units of standard deviation (it is assumed that the standard deviations of the two distributions are equal). The greater the separation, the greater the sensitivity and the greater the d′. In most applications of SDT, d′ ranges between 0.5 and 2.0.

Some signal generation systems may create more noise than others; some people may have more internal noise than others. The greater the amount of noise, the smaller d′ will be. Also, the weaker and less distinct the signal, the smaller d′ will be. Another factor that influences d′ is the ability of the person to remember the physical characteristics of a signal. For example, when memory aids are supplied, sensitivity increase (Wickens, 1992).

Procedures to Calculate SDT Probabilities

a. In SDT, the detection values are expressed as probabilities.

b. The probability of a hit (P_h), miss (P_m), false alarm (P_{fa}), and correct rejection (P_{cr}) are determined by dividing the number of occurrences in a cell (refer to Figure B-1) by the total number of occurrences in a column.

c. The P_h (also referred to as the probability of detection, P_d) will be calculated by dividing the number of threats detected (number of hits) by the total number of hits and misses: $P_m = 1 - P_d$

d. The P_{fa} will be determined by the number of false alarms divided by the total number of false alarms and correct rejections: $P_{cr} = 1 - P_{fa}$

Procedures to Calculate the Response Criterion Measure β

a. Find the false alarm rate from the outcome matrix in the HIT/FA column of Table B-1.

b. Read across the table to the ORD column (for ordinate, the height of the curve).

c. Determine the value table there and write it down.

d. Repeat these operations for the hit rate.

e. Calculate β using the following equation: $\beta = ORDh / ORDfa$.

Procedures to Calculate the Response Criterion Measure c

One recent measure of response bias is c (Ingham, 1970; Macmillan & Creelman, 1990; See, Warm, Dember & Howe, 1995; Snodgrass & Corwin, 1988). The chief difference between the measure c and its parametric alternative β lies in the manner in which they locate the observer's criterion. Whereas the bias index β locates the observer's criterion by the ratio of the ordinates of the signal-plus-noise (SN) and noise (N) distributions, c locates the criterion by its distance from the intersection of the two distributions measured in z-score units. The intersection defines the point where bias is neutral, and location of the criterion at that point yields a c value of 0. Conservative criteria yield positive c values, and risky criteria produce negative c values. See, Warm, Dember, & Howe (1995) conducted three experiments to determine which of five response bias indices (β, c, $\underline{B}''$, $\underline{B}'_H$, and $\underline{B}''_D$) defined by the theory of signal detection provides the most effective measure of the observer's willingness to respond in the context of a vigilance task. The results indicated that c was the most effective of all five indices. The measure c is computed as follows:

$$\underline{c} = .5\,(Z_{fa} + Z_h)$$

Table B-1. Representative Z-Scores and Ordinate Values of the Normal Curve for Different Response Probabilities to Calculate β and D′

HIT/FA	Z	ORD
.01	2.33	0.03
.02	2.05	0.05
.03	1.88	0.07
.04	1.75	0.09
.05	1.64	0.10
.08	1.40	0.15
.10	1.28	0.18
.13	1.13	0.21
.15	1.04	0.23
.18	0.92	0.26
.20	0.84	0.28
.25	0.67	0.32
.30	0.52	0.35
.35	0.38	0.37
.40	0.25	0.39
.45	0.12	0.40
.50	0.00	0.40

HIT/FA	Z	ORD
.50	0.00	0.40
.55	-0.12	0.40
.60	-0.25	0.39
.65	-0.38	0.37
.70	-0.52	0.35
.75	-0.67	0.32
.80	-0.84	0.28
.82	-0.92	0.26
.85	-1.04	0.23
.88	-1.18	0.20
.90	-1.28	0.18
.92	-1.40	0.15
.95	-1.64	0.10
.96	-1.75	0.90
.97	-1.88	0.07
.98	-2.05	0.05
.99	-2.33	0.03

Sensitivity (d') Using the Receiver Operating Characteristic

There are four possible outcomes of a detection task, as shown in Figure B-1. The frequency of the four events can be determined from knowledge of the number of hits and the number of false alarms. So, this pair of values completely specifies the observer's performance. This pair of values can be plotted as one point on a figure known as the Receiver Operating Characteristic (ROC) curve (refer to Figure B-3). The ROC curve shows the hit and false alarm rates possible for a fixed sensitivity (d′) and different response criteria the observer may employ. If each of the two observers (or one observer in two different conditions) behaves according to the theory, you could compare the sensitivity, d′, of the observers (or the conditions) by comparing the ROC curves on which data fall. The nearer the points are to the (0,1) corner, the higher the sensitivity. The nearer the points are to the (0,0) corner, the more conservative the criterion, and the nearer to the (1,1) corner, the more liberal the criterion. Thus, the effectiveness of the operator (or system), as reflected by its ROC curve, is directly proportional to the curve's proximity to the upper left hand corner of the plot in Figure B-3 (Kantowitz & Sorkin, 1983).

In practice, d′ can be determined by comparing the experimentally-determined ROC curve to standard ROC curves (Gescheider, 1976) or by statistical methods (Macmillan & Creelman, 1991).

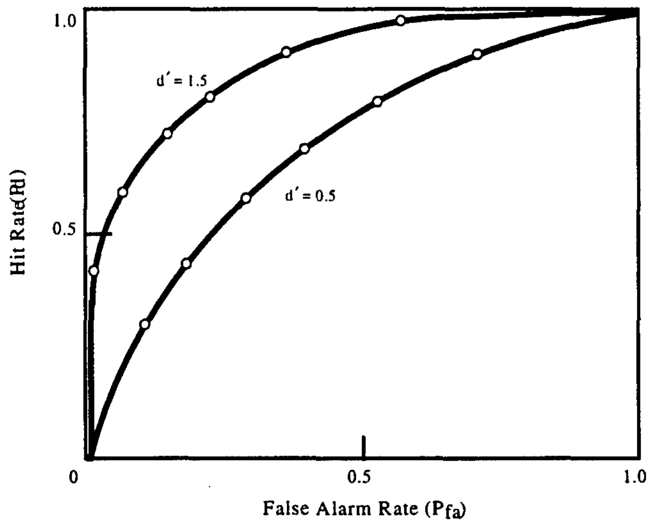

Figure B-3. Sample ROC Curves for Two Observers

Sensitivity (d') Using the Calculation Method

The value of d' can also be determined by calculations based on the proportion of hits and false alarms. This mathematical procedure makes it possible to determine a person's sensitivity based on one data point from a ROC curve (Goldstein, 1984). It should be noted that the calculation method for d′ becomes less reliable as the P_{fa} approaches zero (due to the convergence of ROC curves for $P_{fa} \approx 0.0$ as shown in Figure B-3). The trade-offs between the ROC method and the calculation method are presented in Table B-2. The procedures required to calculate d′ are listed below (Coren & Ward, 1984):

a. Find the false alarm rate from the outcome matrix in the HIT/FA column of Table B-1.

b. Read across the table to the Z column and write it down. This is the Z_{fa}.

c. Repeat these operations for the hit rate, calling the tabled value Z_h.

d. Calculate d′ using the following equation: $d' = Z_{fa} - Z_h$

Table B-2. Trade-Offs Between D′ Methods

	ROC Curve Method	Calculation Method
Benefits	More robust than calculated d′.	Less time required for data acquisition.
Disadvantages	Time requirements for more data to generate ROC points and curves.	Less accurate measure than of d′ than ROC curve method.
Risks	Inappropriate point spread to establish accurate ROC curves.	May be only a preliminary indicator because the measure or data could be too gross.

<u>References</u>

Coren, S., & Ward, L. M. (1984). *Sensation and Perception.* Harcourt Brace Jovanovich, New York.

Foley, P. & Moray, N. (1987). Sensation, Perception, and Systems Design, in Salvendy, G. (ed.) *Handbook of Human Factors.* John Wiley and Sons, New York.

Gescheider, G. A. (1976). *Psychophysics: Method and theory.* Lawrence Erlbaum Associates, Hillsdale, NJ.

Goldstein, E. B. (1984). *Sensation and perception.* Wadsworth Publishing Company, Belmont, California.

Green, D. M., & Swets, J. A. (1988). *Signal detection theory and psychophysics.* John Wiley and Sons, New York.

Ingham, J. G. (1970). Individual differences in signal detection. *Acta Psychologica, 34*, 39-50.

Kantowitz, B. H., & Sorkin, R. D. (1983). *Human factors: Understanding people-system relationships.* John Wiley and Sons, New York.

Macmillan, N. A., & Creelman, C. D. (1990). Response bias: Characteristics of detection theory, threshold theory, and "nonparametric" indexes. *Psychological Bulletin, 107*, 401-413.

Macmillan, N. A., & Creelman, C. D. (1991). *Detection Theory: A user's guide.* Cambridge University Press, Cambridge.

Sanders, M. S., & McCormick, E. J. (1987). *Human factors in engineering and design*, McGraw-Hill Book Company, New York.

See, J. E., Warm, J. S., Dember, W. N., & Howe, S. R. (1995). *Vigilance and signal detection theory: An empirical evaluation of five measures of response bias.* University of Cincinnati, Cincinnati, OH.

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General, 117*, 34-50.

Swets, J. A. (1995). *Signal detection theory and ROC analysis in psychology and diagnostics.* Bolt, Beranek and Newman, Inc. Boston, MA.

Wickens, C. D. (1992). *Engineering psychology and human performance.* Harper-Colins, New York.

APPENDIX C

HUMAN FACTORS DEFICIENCY IMPACT RATING SCALE

Severity	Description
Severe	There is a high probability of operational failure, severe damage, loss of equipment, and injury to operators or passengers.
Major	There is a high probability of degraded system performance, major damage to equipment, or discomfort to operators or passengers.
Moderate	There may be no measurable impact on system performance, though there is a measurable impact upon the performance of system components or sub-systems (including the human subsystem). Operators or passengers try to compensate for, or work around, system defects.
Minimal	There is no measurable impact on the performance of system components or subsystems (including the human subsystem), although operators' or passengers' negative attitudes toward features to the system may be measurable.
Negligible	The problem has a negligible impact on short-term system performance. There is may no measurable impact on operator or passenger attitudes.
None	No problem or negative factor related to system performance is noted.

APPENDIX D

SCREENERS' SURVEY ON TIP USABILITY

SCREENERS' SURVEY ON TIP USABILITY

SUBJECT #____________ DATE_______________

The FAA wants to know how useful this TIP training is and how it can be made better. Your opinion will be important in improving the TIP program. Your responses will be kept secret, so please give ratings that are your honest opinion of the TIP. We thank you for your help.

If you have any questions while taking this survey, please ask the FAA representative for help.

SCREENERS' SURVEY ON TIP USABILITY

SUBJECT #____________ DATE______________

I. The numbered sentences refer to the training that you have just received. Read each sentence and decide whether you agree with it. Use the rating scale to express your agreement or disagreement. The scale goes from 1 to 5. If you circle 5, it means you agree very much with the sentence. If you circle 1, it means you do not agree at all. Use numbers in the middle to express agreement somewhere between the extremes. Circle one number for each sentence.

	Not at all				**Very much**
	1	**2**	**3**	**4**	**5**
TIP improved my ability to find threats in checked bags.	1	2	3	4	5
TIP helped me to do my job.	1	2	3	4	5
The purpose of TIP was explained very well.	1	2	3	4	5
The training on the use of TIP was complete.	1	2	3	4	5
TIP slows down the flow of baggage.	1	2	3	4	5
TIP was easy to use.	1	2	3	4	5
TIP reminds you of the need to stay aware of possible threats.	1	2	3	4	5
TIP helps to teach you to recognize explosive devices.	1	2	3	4	5
I was very tired at the end of the shift because of TIP.	1	2	3	4	5
TIP images were difficult to resolve.	1	2	3	4	5
TIP gave me a good idea about how well I was doing.	1	2	3	4	5
TIP made the job a lot harder to do.	1	2	3	4	5
TIP increases interest in and enjoyment of the job.	1	2	3	4	5

SCREENERS' SURVEY ON TIP USABILITY

SUBJECT #____________ DATE______________

II. Think about the following questions and give the best answer you can think of.

Please list the most important ways in which TIP has affected your job.

1.____

____

2.____

____

3.____

____

4. Did you notice any feature of the TIP images that made it possible for you to determine that it was a TIP besides the presence of an explosive device?

YES NO

5. If you answered yes to the above question, what was different about the TIP images.

____

____

APPENDIX E

TIP USABILITY SUPERVISOR'S SURVEY

The FAA wants to know how useful TIP is and how it can be made better. Your opinion will be important in improving the TIP program. Your responses will be kept secret, so please give ratings that are your honest opinion of the TIP. We thank you for your help.

If you have any questions while taking this survey, please ask the FAA representative for help.

I. The numbered sentences on the next page refer to the training that you have just received. Read each sentence and decide whether you agree with it. Use the rating scale to express your agreement or disagreement. The scale goes from 1 to 5. If you circle 5, it means you agree very much with the sentence. If you circle 1, it means you do not agree at all. Use numbers in the middle to express agreement somewhere between the extremes. Circle one number for each sentence.

	Not at all				**Very much**
	1	2	3	4	5
You can pick the TIP images that you wish to present.	1	2	3	4	5
You can control the rate of TIP presentation.	1	2	3	4	5
You can control the whether TIP provides feedback when a screener misses a threat image.	1	2	3	4	5
You can look up current summary information for particular screeners.	1	2	3	4	5
The purpose of TIP was adequately explained.	1	2	3	4	5
The training on the use of TIP was complete.	1	2	3	4	5
TIP helps you monitor the performance of screener's that you supervise.	1	2	3	4	5
TIP slows down the flow of baggage.	1	2	3	4	5
TIP helps you perform your job.	1	2	3	4	5
TIP increases the vigilance of the people you supervise.	1	2	3	4	5
TIP helps to teach screener's to recognize explosive devices.	1	2	3	4	5
TIP increases screener's interest in and enjoyment of their jobs.	1	2	3	4	5

II. Please list the most important ways in which TIP has affected your job and the jobs of those whom you supervise.

1.____

____

2____

____

3. ____

____

Were you able to pick up any feature of the TIP images that made it possible for you to determine that it was a TIP image without having to look for the explosive device?

YES NO

If you answered yes to the above question, what was the unique feature of the TIP images that you noticed.

____

____

APPENDIX F

PERSONAL INFORMATION FORM

PERSONAL INFORMATION FORM

Please provide some basic information about yourself. All of this information will remain secret. Please do not write your name.

Subject Number: _________ Date: ____ / ____ / 19___

1. At what airport do you work? ____________________________________

2. For what screening company do you work? __________________________

2. What is your gender (please circle your answer)?

   Male Female

3. Have you previously been a certified screener (please circle your answer)?

   Yes No

4. Is English you native or first language (please circle your answer)?

   Yes No

5. Do you wear eyeglasses or corrective lenses (please circle your answer)?

   Yes No

6. Please circle the highest education level that you have completed.

   8th Grade 12th Grade GED Some College College Graduate

7. How much computer experience do you have (please circle your answer)?

   None Very Little Some A Lot

8. Please circle your age:

18-21 22-29 30-39 40-49 50-59 60-69 70 +

APPENDIX G
INFORMED CONSENT.

INFORMED CONSENT

I, ________________________, have received a briefing by the FAA representative about the purpose of these questionnaires. They are designed to find out how screeners feel about the CTX 5000 TIP. I fully understand the purpose of the questionnaires and I have had the opportunity to get information from the FAA representative.

My name and all of my answers to the Personal Information Form and the Screener Survey will be kept strictly CONFIDENTIAL. I will use a code number to keep my identity unknown to my employer and the FAA.

I have been informed that I have the right to quit this test at any time and for any reason. I have also been informed that if any additional details are needed, I may ask one of the administrators present today or I may call Dr. James Fobes at (609) 485-4944 at the end of the test.

Signed: ____

I certify that I am at least 18 years of age.

Date: _____/______/19_____

FAA Witness: ____

Date: _____/______/19_____

APPENDIX H
TIP USABILITY CHECKLIST

Human Factors Principle	Deficiency Rating	Comments
DATA DISPLAY		
1. Ensure that whatever data a user needs for any transaction will be available for display.		
2. Do not overload displays with extraneous data.		
3. For any particular type of data display, maintain consistent format from one display to another.		
4. Ensure that each data display will provide needed context, recapitulating prior data as necessary so that a user does not have to rely on memory to interpret new data.		
5. The wording of displayed data and labels should incorporate familiar terms and the task-oriented jargon of the users.		
6. Choose words carefully and then use them consistently.		
7. When abbreviations are used, choose those abbreviations that are commonly recognized and do not abbreviate words that produce uncommon or ambiguous abbreviations.		
8. Ensure that abbreviations are distinctive so that abbreviations for different words are distinguishable.		
9. When a critical passage merits emphasis to set it apart from other text, highlight that passage by bolding, brightening, color coding, or some auxiliary annotation.		
10. Organize data in some recognizable order to facilitate scanning and assimilation.		

Human Factors Principle	Deficiency Rating	Comments
11. In designing text displays, especially text composed for user guidance, strive for simplicity and clarity of wording.		
12. Use consistent logic in the design of graphic displays and maintain standard format, labeling, etc.		
13. Tailor graphic displays to user needs and provide only those data necessary for user tasks.		
14. When graphics contain outstanding or discrepant features that merit attention by a user, consider displaying supplementary text to emphasize that feature.		
15. When a user's attention must be directed to a portion of a display showing critical or abnormal data, highlight that feature with some distinctive means of coding.		
16. Adopt a consistent organization for the location of various display features from one display to another.		
17. Assign consistent meanings to symbols and other codes, from one display to another.		
18. Choose colors for coding based on conventional associations with particular colors.		

Human Factors Principle	**Deficiency Rating**	**Comments**
SEQUENCE CONTROL		
1. Defer computer processing until an explicit user action has been taken.		
2. Employ similar means to accomplish ends that are similar, from one transaction to the next, from one task to another, throughout the user interface.		
3. Display some continuous indication of current context for reference by the user.		
4. Adopt consistent terminology for online guidance and other messages to users.		
5. Choose names that are semantically congruent with natural usage, especially for paired opposites (e.g., UP / DOWN).		
6. Ensure that the computer acknowledges every entry immediately; for every action by the user there should be some apparent reaction from the computer.		
7. When a user is performing an operation on some selected display item, highlight that item.		
8. Design the interface software to deal appropriately with all possible control entries, correct and incorrect.		
9. When a user completes correction of an error, require the user to take an explicit action to reenter the corrected material; use the same action for reentry that was used for the original entry.		

Human Factors Principle	Deficiency Rating	Comments
10. When a control entry will cause any extensive change in stored data, procedures, and/or system operation, and particularly if that change cannot be easily reversed, notify the user and require confirmation of the action before implementing it. Provide a prompt to confirm actions that will lead to possible data loss.		
USER GUIDANCE		
1. When the computer detects an entry error, display an error message to the user stating what is wrong and what can be done about it.		
2. Make the wording of error messages as specific as possible.		
3. Make error messages brief but informative.		
4. Adopt neutral wording for error messages; do not imply blame to the user or personalize the computer, or attempt to make a message humorous.		
5. The computer should display an error message only after a user has completed an entry.		
6. Provide reference material describing system capabilities and procedures available to users for online display.		
7. In addition to explicit and implicit aids, permit users to obtain further online guidance by requesting HELP.		

Human Factors Principle	Deficiency Rating	Comments
DATA TRANSMISSION		
1. Choose functional wording for terms used in data transmission, including messages, for initiating and controlling message transmission and other forms of data transfer, and for receiving messages.		
2. Design the data transmission procedures to minimize memory load on the user.		
3. Design the data transmission procedures to minimize required user actions.		
DATA PROTECTION		
1. Provide automatic measures to minimize data loss from computer failure.		
2. Protect data from inadvertent loss caused by the actions of other users.		
3. Provide clear and consistent procedures for different types of transactions, particularly those involving data entry, change and deletion, and error correction.		
4. Ensure that the ease of user actions will match desired ends; make frequent or urgent actions easy to take, but make potentially destructive actions sufficiently difficult that they will require extra user attention.		
5. When displayed data are classified for security purposes, include a prominent indication of security classification in each display.		
6. When a user requests LOG OFF, check pending transactions involving data entry/change and, if data loss seems probable, display an appropriate advisory message to the user.		

Human Factors Principle	Deficiency Rating	Comments
VISUAL DISPLAYS		
1. Sufficient contrast shall be provided between displayed information and the display background to ensure that the required information can be perceived by the operator under all expected lighting conditions.		
2. Displays shall be located and designed so that they may be read to the degree of accuracy required by personnel in normal operating or servicing positions without requiring the operator to assume an uncomfortable, awkward, or unsafe position.		
3. Where alphanumeric characters appear on CRT-like displays, the font style shall allow discrimination of similar characters, such as the letter l and the number 1 and the letter z and the number 2.		

APPENDIX I
TIP CUSTOMIZATION CHECKLIST

TIP Customization	Deficiency Rating	Comments / Description
1. The proportion of IED bags and false alarms can be controlled		
2. The rate of TIP presentation can be controlled.		
3. The above functions can be carried out easily by supervisory personnel		

APPENDIX J

TIP FEEDBACK CHECKLIST

TIP Feedback	Deficiency Rating	Comments / Description
1. The screener can receive immediate knowledge of results after responding to a threat object image.		
2. Feedback to screeners can be turned off.		
3. Feedback to the user is consistent in terms of content and format.		
4. Feedback is displayed in a consistent position.		
5. As an element of feedback, review of missed cases can be enabled.		
6. Feedback messages use neutral wording. Messages do not imply blame to the user.		
7. Supervisory personnel can easily change the feedback contingencies		

APPENDIX K

TIP CAPABILITIES CHECKLIST

TIP Capabilities	**Deficiency Rating**	**Comments/Description**
1. Report information is displayed in a directly usable form. Conversions are not needed to use data.		
2. Report format is consistent across reports.		
3. Units of measurement are consistent within and across reports.		
4. Common or unambiguous abbreviations are used in the reports.		
5. Report data are ordered sequentially from left to right and top to bottom.		
6. Related data are grouped together.		
7. Each data field is labeled.		
8. Label wording is consistent.		
9. Label wording is distinctive.		
10. Label location is consistent across reports.		
11. Reports are titled.		
12. Screener reports are generated automatically.		
13. Capability statistics can be summarized by: a. Threats presented and outcomes b. Specific screeners c. Checkpoint activity level d. Time of day		

TIP Capabilities	Deficiency Rating	Comments/Description
14. Screener reports can be generated on demand.		
15. The manufacturer has supplied report documentation.		
16. Summary reports are readily understandable by the users (based on structured interviews)		
17. Performance level reports can be provided to a remotely situated supervisor.		

APPENDIX L

TIP INTEROPERABILITY CHECKLIST

TIP Interoperability	Deficiency Rating	Comments/ Description
1. Reports can be transmitted to remote sites.		
2. Documentation is provided that describes the method by which reports are transmitted to remote sites.		
3. The documented method of report transmission can be verified.		

APPENDIX M
TIP SECURITY ACCESS CHECKLIST

TIP Security Access	Deficiency Rating	Comments/Description
1. The system provides controls to limit access rights.		
2. Access control can be limited to the granularity of a single user.		
3. The system provides controls to limit propagation of sensitive information.		
4. Reports are sensitivity labeled.		
5. Data are protected during transmission.		
6. When data are exported, a sensitivity label is included that accurately represents the level of security of the report.		
7. Procedures are established to control access to reports.		
8. Users are required to enter identification before performing any actions.		
9. The system maintains data to authenticate user identity and level of access.		
10. Authentication data cannot be accessed by any unauthorized user.		

APPENDIX N
TIP MEMORY TEST

Examine each bag as it is presented and decide whether it was presented to you at any time since the TIP began. Use the five point scale, circling the appropriate choice to indicate whether you remember the bag and to give you confidence in your decision. The five scale values are:

1. Definitely new.
2. Probably new.
3. Maybe shown before.
4. Probably shown before.
5. Definitely shown before.

	Definitely New				**Definitely Shown Before**
	1	**2**	**3**	**4**	**5**
Bag 1	**1**	**2**	**3**	**4**	**5**
Bag 2	**1**	**2**	**3**	**4**	**5**
Bag 3	**1**	**2**	**3**	**4**	**5**
Bag 4	**1**	**2**	**3**	**4**	**5**
Bag 5	**1**	**2**	**3**	**4**	**5**
Bag 6	**1**	**2**	**3**	**4**	**5**
Bag 7	**1**	**2**	**3**	**4**	**5**
Bag 8	**1**	**2**	**3**	**4**	**5**
Bag 9	**1**	**2**	**3**	**4**	**5**
Bag 10	**1**	**2**	**3**	**4**	**5**
Bag 11	**1**	**2**	**3**	**4**	**5**
Bag 12	**1**	**2**	**3**	**4**	**5**
Bag 13	**1**	**2**	**3**	**4**	**5**
Bag 14	**1**	**2**	**3**	**4**	**5**
Bag 15	**1**	**2**	**3**	**4**	**5**
Bag 16	**1**	**2**	**3**	**4**	**5**